DOCUMENT RESUME

ED 089 667 IR 000 378

AUTHOR	de Regt, W. F.
TITLE	Thesauri in Educational Documentation; Differences in Use Between Thesauri and Classification Schemes.
INSTITUTION	Council for Cultural Cooperation, Strasbourg (France). Ad Hoc Committee for Educational Documentation and Information.
REPORT NO	DECS-Doc-74-2
PUB DATE	3 Jan 74
NOTE	17p.; See also IR 000 376 through IR 000 380
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	Comparative Analysis; Coordination; *Documentation; *Indexing; Information Processing; *Information Retrieval; *Information Systems; Languages; *Thesauri
IDENTIFIERS	*Eudised; UNESCO

ABSTRACT

This paper on Thesauri in Educational Documentation attempts to show by description and analysis the difference in application between thesauri and classification schemes as indexing languages in documentary information systems, and from this to infer possible future developments. First, the differences between the pre-coordination and post-coordination of terms are examined and a distinction is made between the library and documentary systems for indexing literature. Special attention has been paid to the manner of written and systematic presentation of index languages. Next, short, analytic descriptions of a number of existing thesauri or thesauri in preparation serve to introduce the problem of specificity versus universality in classification systems and thesauri. The Eudised Thesaurus is placed side by side with others for this purpose. From the comparison of methods of construction used by the various thesauri, the conclusion is drawn that an increasing integration between classification systems and thesauri may be expected. (Author)

ED 089667

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IR

Strasbourg, 3 January 1974

DECS/Doc (74) 2

Or. Engl.

COUNCIL FOR CULTURAL CO-OPERATION

AD HOC COMMITTEE FOR EDUCATIONAL DOCUMENTATION AND INFORMATION

EUDISED PROJECT

THESAURI IN EDUCATIONAL DOCUMENTATION

Differences in use between Thesauri and Classification Schemes

by

W. F. de REGT
Head of the Documentation Department
Ministry of Education and Sciences
The Netherlands

IR 000 378

33.630
04.6

CONTENTS

	Page
- SYNOPSIS	1
- INTRODUCTION	1
- PRE- AND POST-COORDINATION IN DOCUMENTARY INFORMATION	2
The library method	2
The documentary method	3
- INDEXING LANGUAGES IN DOCUMENTARY INFORMATION	3
- FROM ENCYCLOPAEDIC TO DOCUMENTARY THESAURUS	4
The lexicological thesaurus	5
The conversion thesaurus	5
The documentary thesaurus	5
- ARRANGEMENT OF A THESAURUS	6
- SETTING OUT THE THESAURUS	7
Grouping in descriptor families	7
Diagrammatic representation of descriptor relationships	7
- THE EUDISED THESAURUS AND OTHERS	8
The macrothesaurus	9
The Eudised Thesaurus	9
The universality of thesauri	10
The Thesaurofacet	10
Unesco's BSO/SRC Project	11
- CONCLUSIONS	12
- BIBLIOGRAPHY	13
- APPENDIX I (An example of an arrow graph)	14
- APPENDIX II (An example of a circle graph)	15

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SYNOPSIS

This paper on Thesauri in Educational Documentation attempts to show by description and analysis the difference in application between thesauri and classification schemes as indexing languages in documentary information systems, and from this to infer possible future developments.

Classification systems and thesauri have not been considered from a theoretical or philosophical point of view, but an attempt has been made to find a documentary approach by studying both thesauri and classification systems in the actual forms in which they are presented.

Firstly the differences between the pre-coordination and post-coordination of terms are examined and a distinction made between the library and documentary systems for indexing literature. The importance of an index language as a controlled language system is subsequently explained.

In addition to the differences in the application of classification schemes and thesauri as index languages, special attention has been paid to the manner of written and systematic presentation of index languages. This stems partly from the history of how the documentary thesaurus came into existence. Its construction and arrangement provide valuable material for the study of the methods of providing easy (visual) reference in the documentary thesaurus.

Next, short, analytic descriptions of a number of existing thesauri or thesauri in preparation serve to introduce the problem of specificity versus universality in classification systems and thesauri.

The Eudised Thesaurus is placed side by side with others for this purpose. From the comparison of methods of contruction used by the various thesauri, the conclusion is drawn that an increasing integration between classification systems and thesauri may be expected.

INTRODUCTION

The thesaurus has made its appearance in libraries and documentation centres and in all probability is there to stay. As is the case with a number of other modern phenomena, the appearance of the thesaurus was not merely accidental.

The speed with which the thesaurus has come to occupy an important place in documentation systems is such as to demonstrate that the thesaurus cannot be regarded as a passing phenomenon, a flash-in-the-pan. On the contrary, it must be seen as a new and efficient aid, the logical result of the more practical approach, which includes post-coordination and the computer, adopted by libraries and documentation centres.

If an attempt is to be made to describe the differences in practice between the new aid and classification systems in order to get a clearer picture of the thesaurus' place among the other instruments, it is necessary first to go into some basic principles.

This paper is not an attempt to consider thesauri and classification systems from a theoretical or philosophical angle. An effort has been made to arrive at a practical, documentary approach by looking at the thesauri and classifications as systematised presentations of indexing languages designed to facilitate the use of literature data in a data retrieval system.

There are two very important ways of approaching the arrangement of data in a retrieval system. It is essential to consider in some detail the difference between pre-coordination and post-coordination.

PRE- AND POST-COORDINATION IN DOCUMENTARY INFORMATION

In the concepts of pre- and post-coordination the term 'coordination' is used to mean 'correlated arrangement'. In documentary information the term pre-coordination is often used to mean 'arranged according to a permanent or virtually constant correlation'.

The term post-coordination is used to indicate that a desired correlation has not been introduced beforehand but comes into being at whatever moment there proves to be a need for it. There are two important applications in respect to documentary information, viz. the coordination of words and of descriptors.

In coordinating words, only pre-coordination is possible. This may be referred to as linguistic pre-coordination. When compound words are made, new words come into being whose component parts lose some or all of their original meaning and the pre-coordinated word acquires an independent meaning.

Pre-coordination of words plays an important part in determining and constructing descriptors. By descriptor is meant a concept in the form of a word, phrase or notation deliberately and explicitly chosen to be used as:

- a representative of the content of the document in a retrieval system (storage side);
- a search feature for retrieving literature data from a retrieval system (retrieval side).

A descriptor may thus be either a single word or a combination of pre-coordinated words. Once it has been designated as a descriptor the compound word acquires a permanent structure.

If a number of different descriptors together indicate the contents of a document, these descriptors may be stored in a retrieval system and retrieved by means of pre-coordination and post-coordination.

These methods will be referred to from now on as the Library Method and the Documentary Method respectively.

The Library Method

A characteristic feature of the conventional library method for indexing literature is that, almost without exception, pre-coordination is employed.

P. Otlet (6) explains it in the following terms: "On entend par classement bibliographique l'art de disposer les ouvrages d'après leur matière (sujet ou contenu) et par classification. La classification bibliographique est une ordre de suite. Elle se développe en une série linéaire unique, dont tous les termes occupent, les un par rapport aux autres, une place ou range désigné par une signe (mots, nombres, ou symboles, quelconques ordonnés en système)".

There is therefore an inherent difficulty in the system which arises when pre-coordinated catalogues are consulted. The user needs to be fully informed of the classification system employed by the cataloguer. To some extent he needs to be a mindreader.

One natural compensation for the rigidity of pre-coordination is cross-referencing or the introduction of any desired permutations. The previously established order of the descriptors is upset. However, it will not often happen that all possible permutations are used.

The Documentary Method

The characteristic feature of this method is the use of post-coordination. The catalogue has almost disappeared and the file or data base has taken its place.

There are three different ways of putting descriptors on file :

- The mono-reference method to one document
(Unit card system). A surrogate document is made (a card or fiche) per descriptor for each document.

- The multi-reference method to one document
(Edge-punched card system). All the descriptors assigned to one document are entered on the information carrier (card etc.) but in such a way that each may be retrieved independently.

- The mono-reference method to a number of documents
(Superimposable punch card system). In this the descriptor assigned to all the documents in a given collection form separate sub-collections with reference to the documents to which they were assigned; a so-called inverted file.

INDEXING LANGUAGES IN DOCUMENTARY INFORMATION

By an indexing language in documentary information is meant the aid which is used in a retrieval system to identify the query and the offer of documentary data. An indexing language differs from a natural language in that it makes use of only a limited number of words or concepts.

In most retrieval systems the original document is represented by a surrogate which gives the content of the relevant document in compressed form. For the purpose of compressing the information an indexing language is used which also expresses systematic and associative correlations.

There are three main reasons for using an indexing language in retrieval systems, which will serve to explain why this method is preferred.

- Reduction in the assignment of the number of concepts to be used ;

- Standarisation of the permitted descriptors ;

- Greater probability of recall.

If an indexing language is presented in writing, showing primarily the systematic connection between the concepts, it is called a classification or code. The natural complement to a classification system or code is an alphabetic list.

If the primary reference to concepts is alphabetical and if in addition the concepts are cross-referenced, the indexing language is called a documentary thesaurus. As one would expect, the natural complement to the thesaurus is the arrangement of descriptors according to their systematic correlation.

A distinction is made, when speaking of indexing languages, between systems with a controlled language and those with a free language. In free language systems there are no constraints in the choice of words to be used in the system.

But this distinction is not an entirely exact one. Even in free language systems the indexing language (aid to the identification of the query or offer of data) is in fact built up in the system. The thesaurus is therefore no longer an independent unit outside the system.

The most striking feature of classification is the hierarchy. Classifications bring attractive advantages to the pre-coordination of descriptors. There is a good deal of similarity between the rules of the hierarchy (whether it is a mono- or poly-hierarchy) and the exclusivity of the linear sequence of library classification. The strict genetic structure of the latter has not proved to be entirely satisfactory and has led to the development of facet classification. Classification may, for that matter, be applied equally efficiently to systems using post-coordination. The thesaurus proves to be an ideal aid in post-coordinating descriptors. The chief reason for this is probably to be found in the fact that post-coordination does not attach exclusive importance to hierarchic relationships. Every relationship, whether hierarchic, associative or linguistic may be important in forming a combination of descriptors at the desired time. Although not so obviously suitable, thesauri can also be applied to systems using pre-coordination.

The differences and similarities in the application of classifications and thesauri are explained more fully in the following pages, with special reference to the arrangement of the thesaurus. Firstly, however, a word about how the documentary thesaurus came into existence will not be amiss.

FROM ENCYCLOPAEDIC TO DOCUMENTARY THESAURUS

The word thesaurus is an ancient one, which in both Greek and Latin means "treasure" or "treasury". In the thirteenth century Brunetto Latini (1220-1294) used it in the sense of an encyclopaedia. Later, in the sixteenth century, it was used to mean a dictionary. Members of the Estienne family of French printers wrote the Thesaurus Linguae Latinae (second edition 1543) and the Thesaurus Graecae (1572).

With the appearance in 1852 of the Thesaurus of English Words and Phrases compiled by Peter Mark Roget, a considerable change in the meaning of the word thesaurus came about.

As we know, a dictionary is an alphabetical list of words, with their meanings, expressed by means of the ideas the words are considered to reflect. Roget's Thesaurus works in exactly the opposite way. He begins with the idea or concept and adds a list of the words which may suitably be used to express it.

Roget's intention was to produce a work which would be of help to writers in finding the right word (or phrase) to describe more or less abstract ideas.

Roget's Thesaurus is divided into three sections:

Part I - The Synopsis of Categories, consisting of six classes further divided into sub-classes and sections;
Part II - The actual thesaurus;
Part III - The alphabetical index.

The remarkable feature is the close similarity between the classification and the thesaurus, and the important role played by classification.

Roget's work has had many imitators; similar publications have appeared in French, Hungarian, Dutch, Swedish, Spanish and German.

The development of the thesaurus idea, however, has also taken other directions. Three different types of thesaurus are discussed below.

The Lexicological Thesaurus

The grouping of the words in a language into interrelated groups is an extremely practical aid to the linguist. One of the best examples of this type of linguistic study is F. Dornseiff's "Der Deutsche Wortschatz nach Sachgruppen", published in 1832.

The object of this work was to group the words of the German language according to their meaning. Dornseiff begins, as Roget was later to do, with general ideas and then investigates as thoroughly as possible the different ways of expressing them in the German language. He lists ordinary conversational language and all derived forms such as underworld slang, popular slang, nursery language, etc.

The work consists of a numbered classification, the main section arranged systematically, and an incomplete alphabetical index. It also contains full linguistic and bibliographical references. Dornseiff is in fact Roget's predecessor. However, the typical "feedback" from phrase to idea, which is so interesting in Roget, is only partly developed. But this was not Dornseiff's main objective.

The Conversion Thesaurus

In a lexicological thesaurus there is hardly any feedback to words, while the conversion thesaurus reverses the process. One no longer begins with ideas, with lists of words attached, but with the words and expressions in common use which are referred back to key words. This method is applied particularly in the study of the possibilities of mechanical translation.

In 1957 (about a century later than Roget) H.P. Luhn developed a conversion thesaurus with the aid of word frequency counts from a given number of documents. He groups synonyms and near-synonyms in families (geni) and each genus is indicated by one keyword. His object is to develop an aid to the mechanical abstraction of texts.

The characteristic features of the conversion thesaurus are therefore the referring back of words or expressions to keywords from a controlled language, and the absence of any classification.

The Documentary Thesaurus

The lexicological and conversion thesauri were developed chiefly for use in the study of linguistic problems. The conversion thesaurus, however, began to enter the field of librarianship and documentation. A further development was to take place in this field.

In 1957, C.L. Bernier and K.F. Heumann published an article (2) entitled "Correlative Indexes". In it they use the term 'thesaurus' to mean a complex network of words with more or less abstract meanings. The least abstract words, expressing things which can be weighed, measured or counted, are used as basic expressions which must be related to as many abstract ideas as possible.

In 1958 T. Joyce and R.M. Needham (4) linked the thesaurus directly with documentary information in their article entitled: "The Thesaurus Approach to Information Retrieval". It was probably Vickery (10) who in 1960 gave a name to the new type of thesaurus and put two of its distinguishing features into words:

1. The documentary thesaurus assists in information storage and retrieval by converting words from everyday language into descriptors;

2. The documentary thesaurus is also a summary of the relationships between the descriptors.

It is now possible to define the documentary thesaurus as follows:

A (documentary) thesaurus is a fixed or chosen list of descriptors and non-descriptors in alphabetical order for use in an information storage and retrieval system, in which hierarchic, associative or linguistic relationships between the descriptors can be shown.

ARRANGEMENT OF A THESAURUS (7)

Descriptors are chosen or fixed. There must therefore always be an authority to select, limit, establish and make rules for the admission of new descriptors.

The descriptors for a retrieval system can be selected in three ways:

- By a process of deduction, motivated by a desire for completeness;
- By a process of induction, for example from frequency counts based on a limited collection of documents;
- From experience, based on the more or less known usefulness of descriptors.

The limitations of a thesaurus may consist in its being restricted to:

- One subject or branch of science;
- One organisation or institution;
- One collection of documents.

Linguistic forms play their part in the fixing of descriptors. A descriptor may be made up of one single word or of a group of pre-coordinated words. The single word is called a uniterm, the pre-coordinated descriptors are called polyterms or multiterms (8).

There is a connection between the specificity and the degree of pre-coordination of a descriptor. Its usefulness will be optimal if on the one hand the frequency of use is not too high as a result of its meaning being too general and on the other hand not too low because it is too specific. The nature of the documents and the query situation of the users of the retrieval system both have some influence on this.

The arrangement of the thesaurus itself is governed by the measure of search guidance it is desired to build into it. The measure of search guidance is determined by the measure in which the correlation between the descriptors is made known. A distinction must be made between natural and fixed relationships.

Among the natural relationships are those between descriptors and non-descriptors. Non-descriptors are words (or word combinations) which cannot be used as search features. Synonyms (different words with the same meaning) are considered as non-descriptors and must be marked ('see', or 'use' ...). Quasi-synonyms (e.g. acronyms) are also non-descriptors and must be back-referenced. Homonyms (words of the same form as others but of different sense) may also be back-referenced. However the meaning is often narrowed down by means of an identifier.

Fixed relationships may be divided into two categories, vertical and horizontal. Vertical relationships are those between genus and species, and horizontal, the associative relationships ('see also').

In a thesaurus it is possible to indicate relationships between descriptors in many ways by means of symbols. The following list is in almost general use, however.

Symbol	Standing for	Meaning
D	Descriptor	The actual term used in search
SN	Scope note	Short explanation of the term
UF	Used for	Summary of the non-descriptors (not for use in system)
BT	Broader term	Genus relationship
NT	Narrower term	Species relationship
RT	Related term	Associative relationship

Looking at the use of the thesaurus from the point of view of a search for literature, the network of the relationships between descriptors appears as a complex, built-in search guide. It is possible to arrange this complicated network in a variety of ways.

SETTING OUT THE THESAURUS

It is possible to show the correlation between descriptors in a thesaurus. There are two distinct methods:

- By grouping descriptors according to their family relationship;
- By diagrams of the relationships between descriptors.

One purpose of making the thesaurus readable is to facilitate its lexicological management. Another reason is to make it easier to find descriptors for the analysis of documents and the search for literature.

This "readability" makes it possible for the user to search systematically for an idea. It must be remembered, however, that the visual system is subordinate to the primary, alphabetical use of the thesaurus and therefore can never be as valuable as a traditional classification system (3).

Grouping in Descriptor Families

In composing descriptor families it must be remembered that each descriptor may belong to one family only. The usual method of grouping is as follows: a scope note is worked out for each distinct family and a separate synopsis is given of all family scope notes. The descriptors belonging to each family are listed alphabetically.

In the thesaurus proper, the main descriptors (i.e. the descriptors for which no BT is to be found in the thesaurus) are followed by a reference to their family relationship. This can be in the form of a notation (numbers and/or letters). The notation is the link between the alphabetical and the systematic parts of the thesaurus. In this way it has also become possible to determine the family relationships of a descriptor in the alphabetical section very quickly.

Diagrammatic representation of descriptor relationships

One of the fundamental problems in documentary information is the identification of the following:

1. the ideas of the author of a document, of the person processing the document and of the person requesting literature, with

The question may now be asked to what extent thesauri, built up as they are from empiricism, may be endowed with some measure of generality, of universality, It does not seem possible to give a definite answer to this question as yet. A number of interesting and in this connection relevant forms of thesaurus have appeared which may perhaps provide some indications as to how the question may be answered.

The Macrothesaurus

The OECD Macrothesaurus (15) is an attempt to demonstrate that it is possible to create a common descriptor language. Its purpose is to coordinate the thesauri already in use by the individual organisations of the UN family. A link mechanism was designed for the various data bases which are each arranged in accordance with their own thesauri. The link mechanism does not contain all the descriptors, neither does it contain only those from all the linked retrieval systems. A selection had to be made in order to ensure that:

- The descriptors would mean the same thing to all users;
- The descriptors would actually be used by a reasonable number of the users;

If the idea of a macrothesaurus is the right one, the structure would have to be such as to leave room for subjects not already included.

It was decided in 1969 to compile the macrothesaurus and this was carried out following strictly inductive methods. After inventorying, seven thousand suggested descriptors were selected and circulated to all the participants in the project with the questions:

- Whether the suggested descriptors were used by that institution, and
- Whether the suggested descriptors should or should not be included.

In this way a general and common collection of descriptors was formed, and the Macrothesaurus is therefore a thesaurus of macro-descriptors which are so general as to be useful to all the participants. The project was limited mainly to aspects of economic development work.

The Eudised Thesaurus

The Eudised Thesaurus (12), which is being developed by the Council of Europe for use in the exchange between member countries of educational information, is being compiled on a different principle from that underlying the macrothesaurus.

The Eudised Thesaurus leaves the national educational thesauri of each participating country intact in principle and is an attempt to gain a multi-lingual entry to the existing educational thesauri. The multi-lingual aspect of the Eudised Thesaurus is a fundamental feature of it and does away with the objections to the use of one common or carrier language.

A carrier language means double work for other language areas when assigning descriptors. In the analysis of documents the carrier language also tends to dominate the user language, which inevitably leads to loss of information.

The subject of education is well known for its strongly national characteristics. Its organisation is often determined by history, and educational ideals are closely linked to the cultural background of the nation. This produces a situation in which national educational thesauri contain descriptors for which there is no equivalent in another language.

2. the same ideas transformed into concepts expressed in ordinary language, and in turn, with

3. the same concepts transformed from ordinary language into descriptors which are part of a controlled language system.

Diagrams of the relations between descriptors can prove extremely helpful for the purpose of such identification. They can demonstrate relationships in a way which neither classifications nor thesauri can.

The arrow and circle graphs are good examples of diagrams showing descriptor relationships.

Constructing arrow graphs

Arrow graphs (8) are constructed by first listing all the terms used for a particular subject. They are subsequently grouped in a number of families of related terms.

The terms belonging to one family are written down in such a way that each is positioned in accordance with its relationship to the other terms.

Once the synonyms and quasi-synonyms have been eliminated, it is possible to construct polygons around terms which are closely related. Then a nucleus term is shown inside each polygon and this term is finally given the status of a descr

If the sides of the polygon are now replaced by arrows, an arrow graph is formed. A single arrow is used to show an NT relationship and a double one for an RT relationship. (See Appendix I.)

Constructing circle graphs

One disadvantage of the arrow graph is that it is not always immediately obvious which is the NT and which the BT in a hierarchic relationship. The circle graph (9) provides a simple solution to this problem. It consists of concentric circles showing the hierarchic levels. (See Appendix II.)

The BTs are found in the centre and derived NTs on the circumference. Arrows are also used in this type of graph, pointing from BT to NT. But beginning with the outer circle, the BT can be found via the NT.

THE EUDISED THESAURUS AND OTHERS (13)

If one regards classifications and thesauri as being confined to controlled language systems, they are essentially vocabularies whose scope is limited. Their scope is largely determined by the purpose for which they are designed.

Classifications, and in particular the universal classifications, have yet another dimension. They claim to cover human knowledge in a particular field (the universal classifications cover the sum total of human knowledge), in a systematic, correlated form. A good example of the universal classification is the Universal Decimal Classification (UDC).

Too close an approach to universality may create difficulties in the direct use of classification in a controlled language system. The almost inexhaustible possibilities of the UDC for constructing detailed concepts by assigning facets to the main concepts means that a second step is urgently needed. A choice will have to be made and a limited collection of concepts compiled, which may be used to the exclusion of all other possiblities. Only in this limited sense can the universal classification be called an index language.

The Eudised Thesaurus might be described as a national thesaurus in the first place, with a common link mechanism for the purpose of international consultation and exchange of information.

The Eudised Thesaurus will therefore consist of at least the following two main sections:

- The systematic synopsis of the descriptor groups for which equivalents exist in the languages of all the participants in the project;

- The alphabetical, classified list of all the descriptors in one language.

The chief problem facing the Eudised Thesaurus is thus the compilation of the first section.

The Universality of Thesauri

Both the Macrothesaurus and the Eudised Thesaurus can claim a certain generality. It has not, however, been proved that the construction methods followed will lead to universality if extended.

The classification school of thought maintains that it is not possible to arrive at a universal thesaurus by way of the construction of the thesaurus.

As we know, the relationships between the descriptors in a thesaurus are arbitrarily fixed. True universality can only be achieved by obeying rules which apply generally. The nature of the thesaurus itself makes it impossible to make any such rules. The territory of a thesaurus may be either discipline-oriented or mission-oriented.

If the territory is discipline-oriented, it would not appear possible for it to go beyond the boundaries of that subject of its own accord. On the other hand, in the thesaurus which is mission-oriented there is scope for a multi-disciplinary approach. However, the ordering of the poly-hierarchy thus obtained would seem to be reserved for a classification system.

A universal superstructure in the nature of classification may well be the solution to the greater universality of a thesaurus. And this is the direction in which the thesaurofacet (1) seems to be pointing.

The UNESCO UNISIST-BSO project presents similar problems.

The Thesaurofacet

The Thesaurofacet is a good example of how the classification system and the thesaurus may be integrated, being as it is both thesaurus and a facet classification system for all engineering and related subjects.

The Thesaurofacet covers the whole field of the natural sciences and technology and consists of two complementary parts, a classified section and a thesaurus.

The classification groups the main subjects in relation to each other, in accordance with the traditional rules of classification. Within each subject, the ideas are arranged in facets; within the facets, in hierarchies and series, with a letter and number notation.

The thesaurus section is arranged according to the usual pattern. BTs and NTs refer to hierarchically broader or narrower terms which do not appear in the part of the classified section which coincides with the descriptor. Non-hierarchic relationships are indicated by "related terms".

There are a great many categories of non-hierarchic relationships which frequently appear in totally different subject areas. Relationships of this kind are not at all easy to classify for convenient reference, but of course they appear in the thesaurus section. The latter section in the Thesaurofacet to some extent replaces the alphabetical index usually appended to a conventional system of classification.

The information which is given in the thesaurus section to go with the descriptor is complementary to the information given by the classified section. Although the Thesaurofacet is primarily constructed as a controlled language, with descriptors for use in controlled indexing and searching, the integration of classification and thesaurus gives the system a much more universal character.

UNESCO's BSO/SRC Project

UNESCO's BSO/SRC Project may serve as a final example to illustrate the relationships between thesauri and classification systems.

In the Unisist Project, involving a world-embracing network of information systems, the need became evident for a switching mechanism which would link up the individual information systems, each with their index language (thesauri, classification, keywords, etc.).

The designing of the mechanism was entrusted to the FID. The project was described as follows during the course of the 1972 congress in Budapest:

> "The need for the establishment of a broad classification scheme or a Broad System of Ordering (BSO) in the framework of the Unisist programme arises naturally as one of the aspects in the process of achieving compatibility between existing and future information systems. Having in mind the great diversity of classification schemes used in the world, it can be stated that the main function of BSO is to serve as a switching mechanism to link different individual classifications and thesauri, in the process of information-transfer."

An FID working group on classification research defines classification as follows:

> "By classification is meant any method for recognising relations, generic or other, between items of information regardless of the degree of hierarchy used and of whether these methods are applied in connection with traditional or computerised information systems."

A special working group of the FID was charged with the building up of a Subject-Field Reference Code (SRC) which would provide Unisist with such a Broad Scheme of Ordering. The working group decided to publish the results of their work in three parts: a scheme of classification with notations, a thesaurus with synonyms, and a graph.

CONCLUSIONS

The rather stormy passage the documentary thesaurus has had on the way to becoming an efficient instrument in librarianship and documentation, and new ideas on the basic principles of classification, may be an indication that thesauri and classification schemes will continue to exercise an influence on each other.

The method of post-coordination makes it obligatory to have full permutation of the descriptors on file. The computer makes this possible and it looks as if this will lead to a more intensive study of the properties of documentary index languages. In constructing a thesaurus one is faced with the problem of achieving universality. The Thesaurofacet provides the first solution to the problem of how the thesaurus and the classification scheme may be integrated. And, finally, the Eudised Thesaurus is a first attempt to find a rational solution to the problem of multi-lingual educational documentation.

BIBLIOGRAPHY

1. Aitchison, J., A. Gomersall and R. Ireland.-
Thesaurofacet; a thesaurus and faceted classification for engineering and related subjects.
Whetstone, 1969. The English Electric Company, Ltd.

2. Bernier, C.L. and K.F. Heumann.-
Correlative indexes. American Documentation.
8 (1957) 211 - 20.

3. Goodman, F.-
The role and function of the thesaurus in education.
In: Thesaurus of Eric descriptors.
New York, 1970. CCM Information Corporation.

4. Joyce, T. and R.M. Needham.-
The thesaurus approach to information retrieval.
American Documentation, 9 (1958) 192 - 7.

5. Macrothesaurus, a basic list of economic and social development terms.-
Paris, 1972. Organisation for Economic Co-operation and Development.

6. Otlet, P.-
Traité de Documentation ; le livre sur le livre.
Editiones Mundaneum, 1934.

7. Richtlinien für die Erstellung und Weiterentwicklung deutschsprachiger Thesauri.-
DIN 1463, 1972. Fachnormenausschluss Bibliotheks- und Dokumentationswesen im Deutschen Normenausschluss (DNA).

8. Rolling, L.-
The role of graphic display of concept relationships in indexing and retrieval vocabularies.
Luxembourg, 1965. European Atomic Energy Community.

9. TDGK-Circular Thesaurus System. 4th Ed.-
The Hague, 1966. Netherlands Armed Forces Scientific and Technical Documentation Centre.

10. Vickery, B.C. Thesaurus a new word in documentation.-
Journal of Documentation, 16 (1960) 181 - 9.

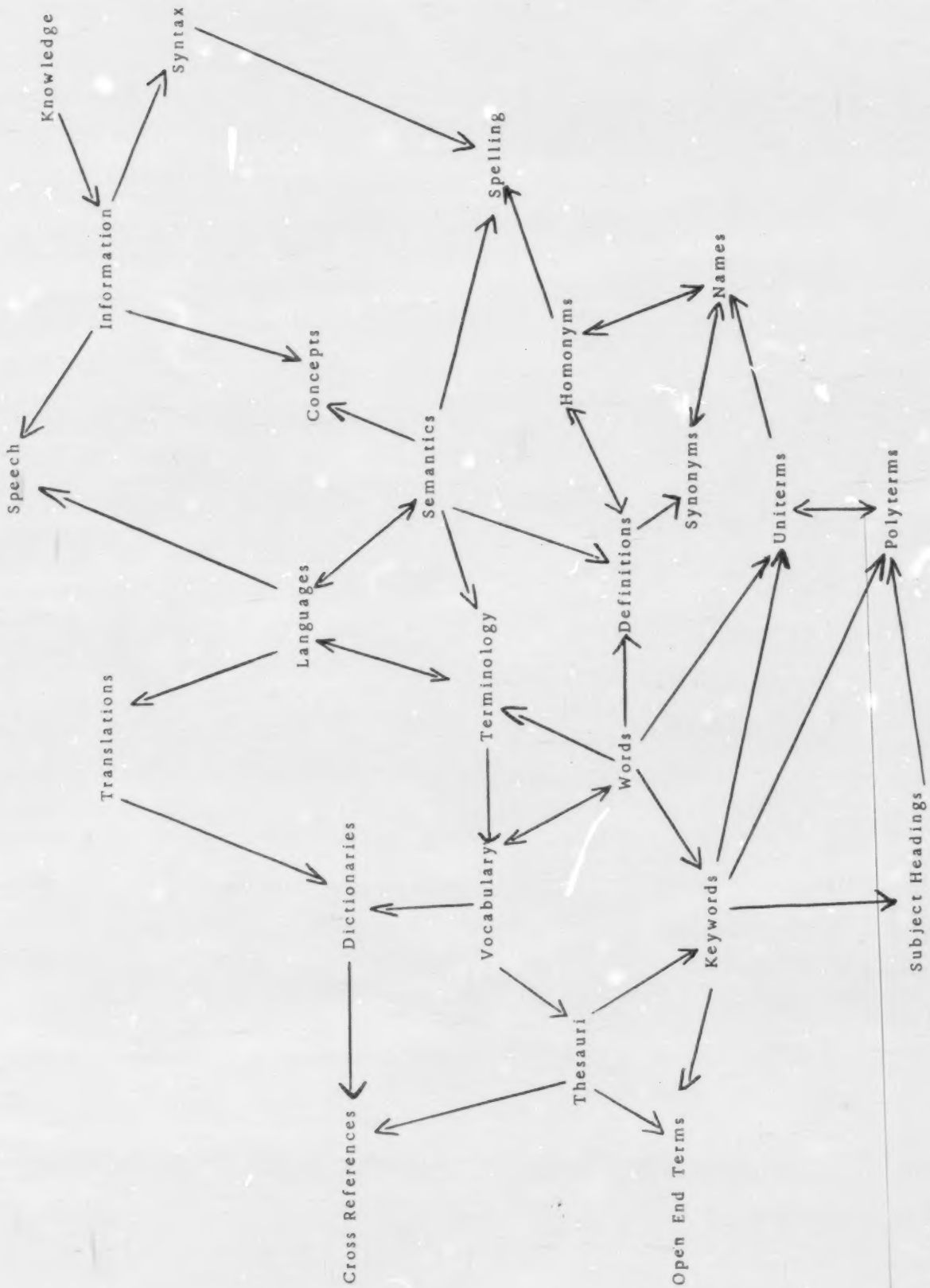
11. Vickery, B.C.-
On retrieval system theory.
London, 1961.

12. Viet, J.-
Problems in compiling the Multilingual Eudised Thesaurus.
Eudised technical studies.
Council of Europe, 1971.

13. Walkley, J. and B. Hay.-
An annotated list of Thesauri held in the Aslib library.
Aslib Proceedings, 23 (1971) 292 - 300.

APPENDIX I

An example of an arrow graph



APPENDIX II

An example of a circle graph